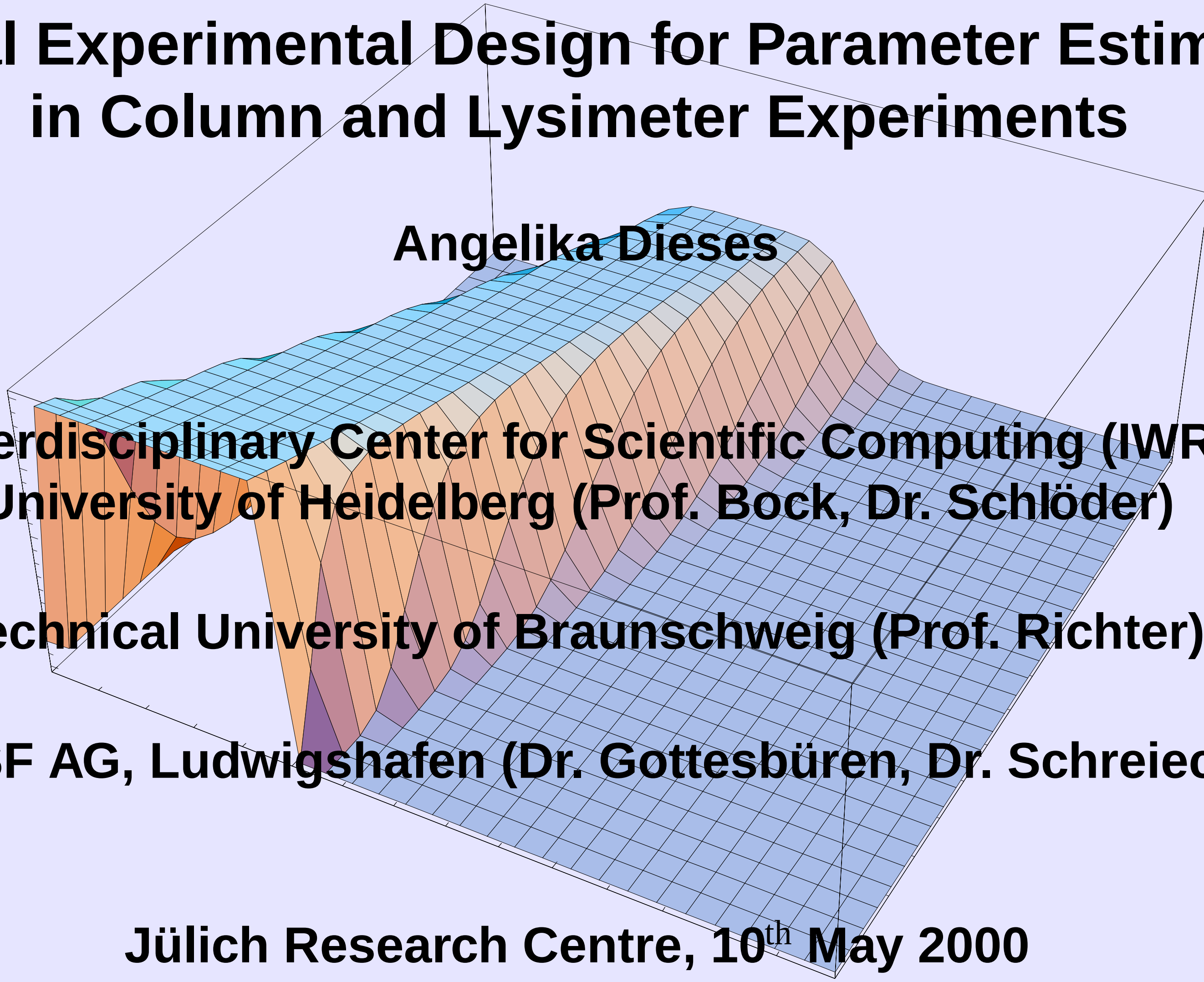
# Optimal Experimental Design for Parameter Estimation in Column and Lysimeter Experiments



**Angelika Dieses**

**Interdisciplinary Center for Scientific Computing (IWR)**
**University of Heidelberg (Prof. Bock, Dr. Schlöder)**

**Technical University of Braunschweig (Prof. Richter)**

**BASF AG, Ludwigshafen (Dr. Gottesbüren, Dr. Schreieck)**

**Jülich Research Centre, 10th May 2000**

# Contents

Why optimal experimental design?

## Numerical Methods

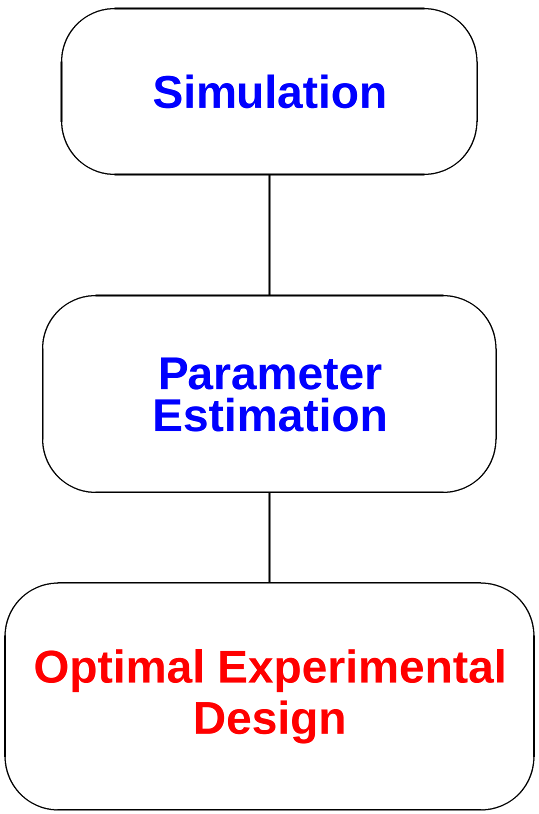
- Formulation of the inverse problem and its efficient solution
- Statistical quality of parameter estimates
- Formulation of the design problem as an optimal control problem
- Numerical solution of the optimization problem

## Application: Design of Column Outflow Experiments

- Scenario 1: Optimization of experimental conditions
- Scenario 2: Simultaneous optimization of experimental conditions and sampling scheme

# Why Optimal Experimental Design?

**AIM: Reliable Simulation Models**




## Motivation

- For registration: Performance of many expensive column and lysimeter studies
- Substitution by computer simulations
  $\Longrightarrow$ enormous savings
- Determination of unknown parameters

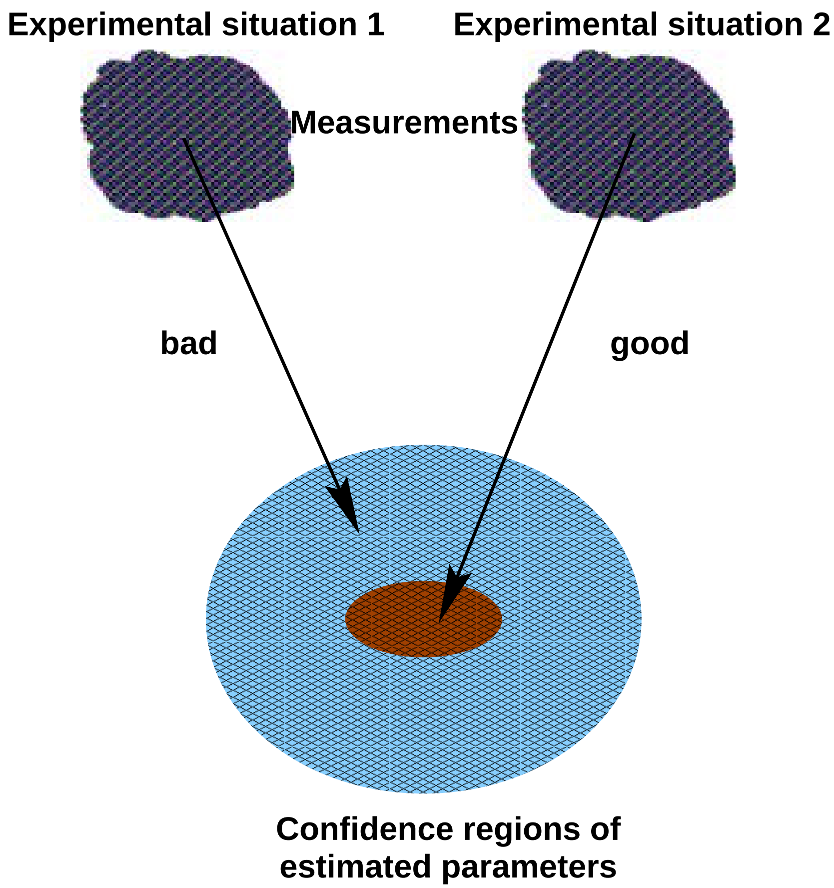
## Problem

- Often unsuitable data for parameter estimation
  $\Longrightarrow$ ill-conditioned or singular problems

## Remedy: “optimal” data

- optimal experimental conditions
- optimal sampling schemes

# Optimal Experimental Design




- How inflow rates and/or input substance concentrations should be chosen?
- What, when and where measurements should be carried out?

# Formulation of the Inverse Problem

**Given**: experimental data $\eta_{kij}$, $(k=\psi,\theta,c_l)$, $(l=1,\ldots,n)$, $(i=1,\ldots,m_1)$, $(j=1,\ldots,m_2)$

$$\eta_{kij} = b_k\big(t_i, z_j, k(t_i,z_j), p\big) + \varepsilon_{kij}, \qquad \varepsilon_{kij} \sim \mathcal{N}\,(0, \sigma^2_{kij})$$

**Aim:** vector of parameters $p$ and a solution $k(t,z)$, such that

$$\begin{aligned}
\min \|F(\psi,\theta,c_l;p)\|_2^2 &= \min \sum_{k=\psi,\theta,c_l} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \sigma_{kij}^{-2}\big(\eta_{kij} - b_k(t_i,z_j,k(t_i,z_j),p)\big)^2 \\
C(\psi;p)\frac{\partial\psi}{\partial t} &= \frac{\partial}{\partial z}\Big(K(\psi;p)\frac{\partial}{\partial z}(\psi - z)\Big) + S_1(\psi;p) \\
\frac{\partial\theta}{\partial t} &= \frac{\partial}{\partial z}\Big(\bar{D}(\theta;p)\frac{\partial\theta}{\partial z} - \bar{K}(\theta;p)\Big) + S_2(\theta;p) \\
\frac{\partial(\theta c_l)}{\partial t} + \frac{\partial(\rho S_l)}{\partial t} &= \frac{\partial}{\partial z}\Big(\theta D_{h_l}(\theta;p)\frac{\partial c_l}{\partial z}\Big) - \frac{\partial}{\partial z}\big(q(p)c_l\big) + Q_l(c_1,\ldots,c_n;p) \\
&+ \text{ initial and boundary conditions}
\end{aligned}$$

# Efficient Solution of the Parameter Estimation Problem

- Discretization: Method of Lines
  - Space: Finite Differences
  - Time: Multiple Shooting

$\Longrightarrow$ large scale, nonlinear, constrained least-squares problem

$$\min_p \|r_1(y(t_0;p)\ldots,y(t_f;p),p)\|_2^2 = \min_p \sum_{i,j} \frac{(\eta_{ij} - b_i(t_j,y(t_j;p),p)^2}{\sigma_{ij}^2}$$
$$r_2(y(t_0;p),\ldots,y(t_f;p),p) = 0$$

- Reduced generalized Gauss Newton method
  - very efficient generation of derivatives

$\Longrightarrow$ **ECOFIT** (Dieses, Schlöder, Bock)

# Statistical Quality of Parameter Estimates

Variance-Covariance Matrix

$$C(q,w,\hat{p}) = J(q,w,\hat{p})^{+} \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} J(q,w,\hat{p})^{+T}$$

- Experimental conditions: control functions $q$
- Sampling scheme: weights for measurement points $w$

$\Longrightarrow$ Parameter estimation: $q$ and $w$ fixed

$\Longrightarrow$ Optimal experimental design: $\hat{p}$ fixed

$$C(q,w,\hat{p}) = J(q,w,\hat{p})^{+} \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix} J(q,w,\hat{p})^{+T}$$

# Optimal Experimental Design Problem

1. Choose an optimality criterion (A-,D-,E-Optimality)
   here: A-Optimality: average variances of parameter estimates

$$\Phi_1(C) = \frac{1}{n} trace(C)$$

2. Define
   (a) Set of feasible experimental conditions $E$
   (b) Set of $n$ feasible measurement points $M$

3. Determine
   (a) Optimal experimental conditions $q^* \in E$
   (b) Optimal sampling scheme, e.g. weights $w^* \in \{0,1\}^n$

   with

$$\Phi_1(C(q^*, w^*)) = \min_{q \in E, w \in \{0,1\}^n} \Phi_1(C(q, w))$$

# Nonlinear State-Constrained Optimal Control Problem

$$\min_{q,w} \Phi\Big(C\big(J(y(p,q),p,q,w)\big)\Big)$$
$$\dot{y} = f(t,y,p,q)$$

Control and state constraints:

$$c(y(t),q) \left\{ \begin{array}{c} = \\ \geq \end{array} \right\} 0$$

Interior-point conditions:

$$d(y(t_0;p,q,w)\ldots,y(t_f;p,q),p,q,w) \left\{ \begin{array}{c} = \\ \geq \end{array} \right\} 0$$

**Difficulty:** Objective function that implicitly depends on the Jacobian of the underlying parameter estimation problem in PDEs and ODEs.

# Numerical Solution of the Optimization Problem

- **Direct Approach:** Parameterization of control functions (piecewise constant/linear)

$\Longrightarrow$ finite dimensional nonlinear constrained optimization problem

$$\min_{\nu} F(\nu)$$
$$G(\nu) = 0$$
$$H(\nu) \geq 0$$

$\nu$: Parameterization of controls $q$ and weights $w$.
$\Longrightarrow$ solved by structured SQP-method

$\Longrightarrow$ **VPLAN** (Körkel, Bauer, Bock, Schlöder, Sager, Rücker)

$\Longrightarrow$ **ECOPLAN** (Dieses, Körkel, Schlöder, Bock)

# Typical Column Outflow Experiment

**Water transport:** $\Longrightarrow$ **3 unknown parameters** $n$, $\alpha$, $K_s$

$$C(\psi)\frac{\partial\psi}{\partial t} = \frac{\partial}{\partial z}\Big(K(\psi)\frac{\partial}{\partial z}(\psi - z)\Big)$$

$$K(\psi) = K_s\frac{(1-(\alpha|\psi|)^{n-1}(1+(\alpha|\psi|^n)^{1/n-1})^2}{(1+(\alpha|\psi|)^n)^{\frac{1-1/n}{2}}}$$
$$C(\psi) = \alpha(n-1)(\theta_s-\theta_r)(\alpha|\psi|)^{n-1}(1+(\alpha|\psi|)^n)^{1/n-2}$$

- Initial condition: $\psi(0,z) = 670, \qquad z \geq 0$
- Upper boundary condition:

$$q(t,0) = -K(\psi(t,0))\Big(\frac{\partial\psi(t,0)}{\partial z} - 1\Big) \quad \textit{control function}$$

- Lower boundary condition: $\frac{\partial\psi(t,L_e)}{\partial z} = 0, \qquad t \geq 0$

**Solute transport:** $\Longrightarrow$ **3 unknown parameters $k$, $b$, $D_m$**

$$\frac{\partial(\theta c)}{\partial t} = \frac{\partial\left(\theta D_h(\theta)\frac{\partial c}{\partial z}\right)}{\partial z} - \frac{\partial(qc)}{\partial z} - kc$$

$$D_h(\theta) = \frac{0.0046e^{b\theta}}{\theta} + D_m$$

- Initial condition: $c(0,z) = 0, \qquad z \geq 0$
- Upper boundary condition:

$$-D_h(\theta(t,0))\frac{\partial c(t,0)}{\partial z} + \frac{q(t,0)}{\theta(t,0)}c(t,0) = \frac{q(t,0)}{\theta(t,0)}c_0(t,0)$$

- Lower boundary condition: $\frac{\partial c(t,L_e)}{\partial z} = 0, \qquad t \geq 0$

$\Longrightarrow$ Coupled water and solute transport: 6 unknown parameters

# Scenario 1: Optimization of Experimental Conditions

## Experimental Conditions: 2 Control functions

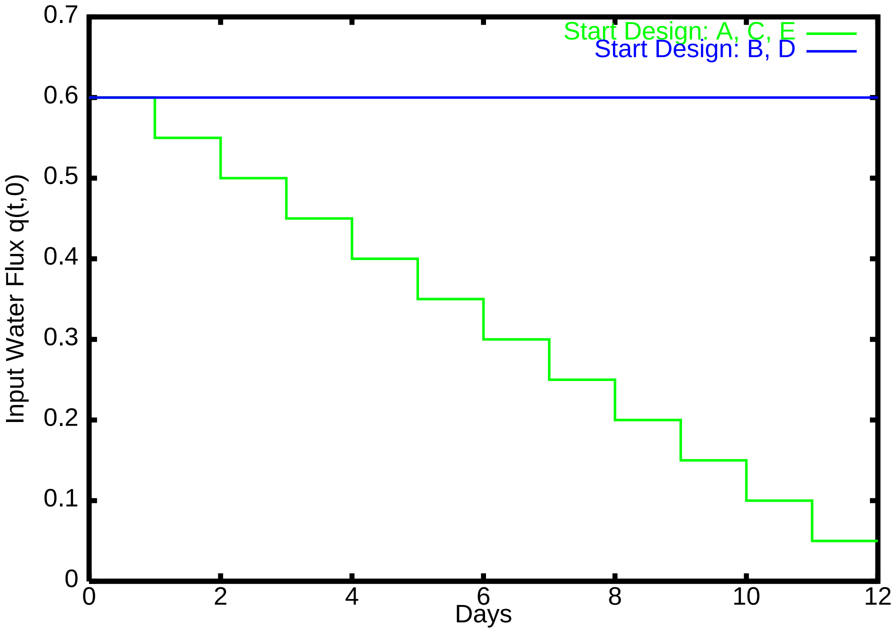
- Parameterization of the input water flux $q(t,0)$
  - Changing in controls possible every 24 hours
  - Piecewise constant in the range of [0, 0.6]
- Parameterization of the substance input concentration $c_0(t,0)$
  - Changing in controls possible every 24 hours
  - Piecewise constant in the range of [50,200]

## Sampling Scheme: Fixed Measurements

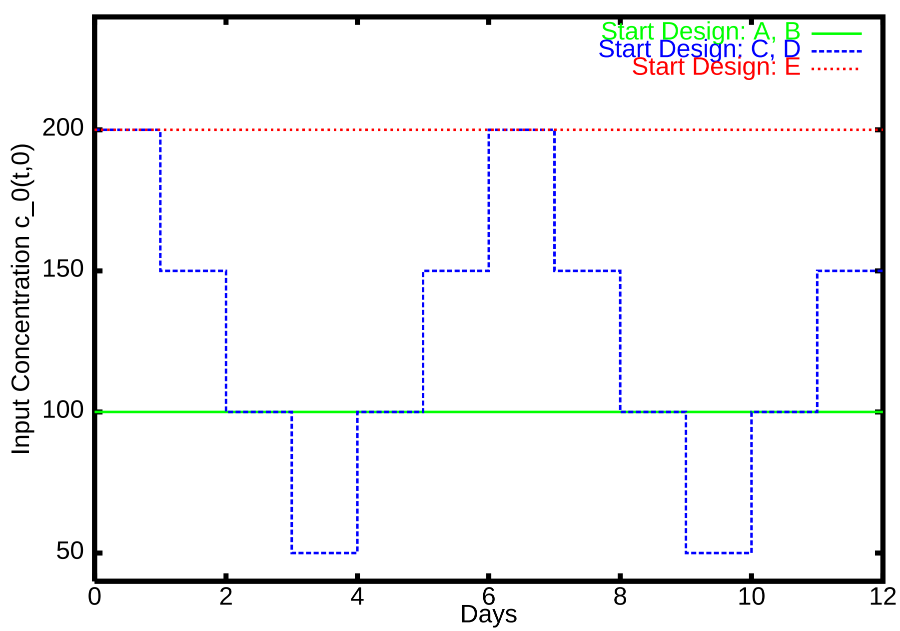
- 24 outflow measurements within 12 days: every 12 hours ($\sigma = 0.01$)
- 6 profile measurements ($\sigma = 0.1$)
  - At the end of the experiment ($T$=12 days)
  - Determination of concentration $c$ in 6 depths.

# Parameterized Start Designs

## Input water flux $q(t,0)$




## Substance input concentration $c_0(t,0)$



# Quality of Start Designs: A-Criterion Values and Standard Deviations

## Start Design A

A-Criterion = 3.8859

| | scaled | | standard deviation |
|---|---|---|---|
| $n$ | 1.0 | $\pm$ | 0.1604 |
| $\alpha$ | 1.0 | $\pm$ | 1.5968 |
| $K_s$ | 1.0 | $\pm$ | 3.3752 |
| $k$ | 1.0 | $\pm$ | 0.0124 |
| $b$ | 1.0 | $\pm$ | 3.0460 |
| $D_m$ | 1.0 | $\pm$ | 0.2644 |

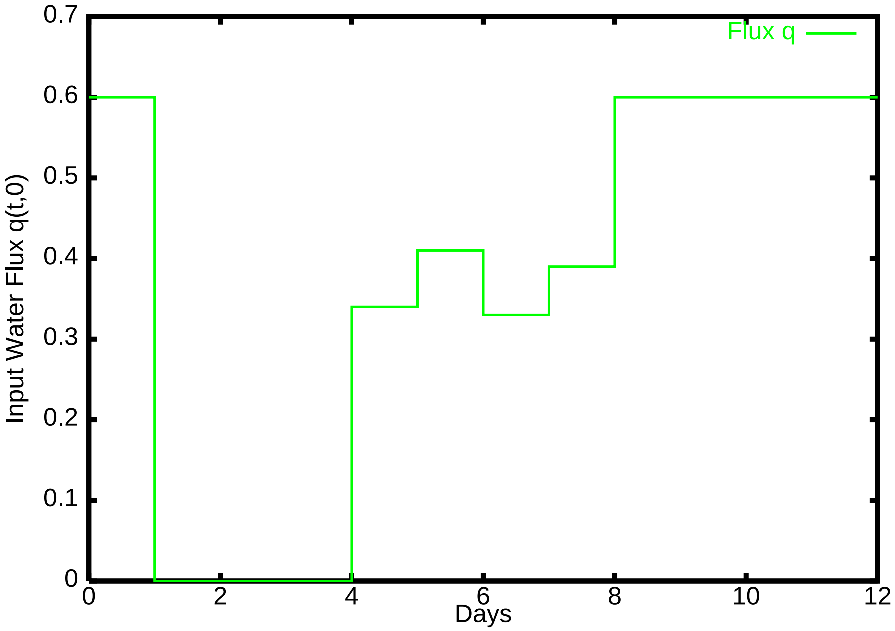
## Start Design B

A-Criterion = 0.75

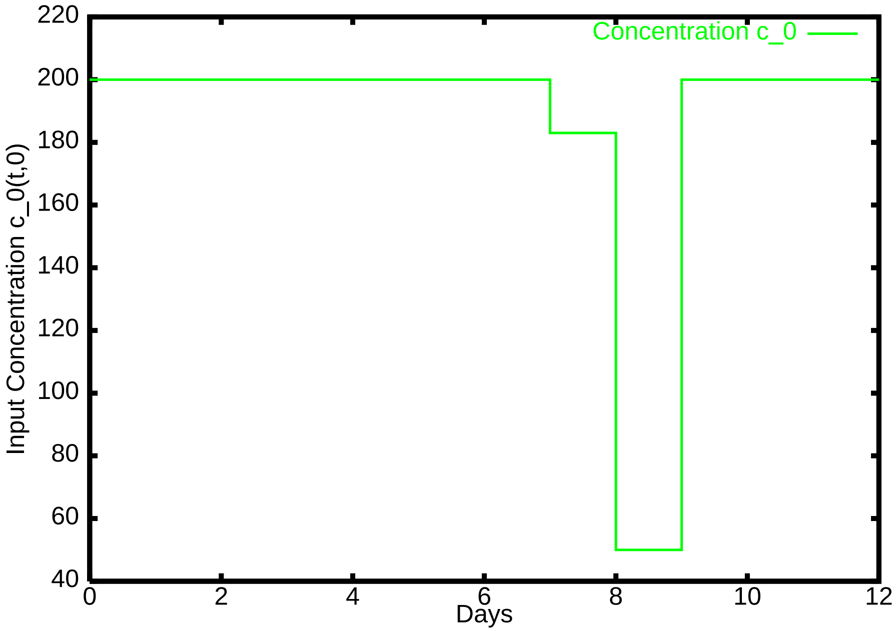
| | scaled | | standard deviation |
|---|---|---|---|
| $n$ | 1.0 | $\pm$ | 0.0800 |
| $\alpha$ | 1.0 | $\pm$ | 0.8600 |
| $K_s$ | 1.0 | $\pm$ | 1.6736 |
| $k$ | 1.0 | $\pm$ | 0.0083 |
| $b$ | 1.0 | $\pm$ | 0.9712 |
| $D_m$ | 1.0 | $\pm$ | 0.1033 |

# Optimal Experimental Conditions

**Input water flux** $q(t,0)$




**Substance input concentration** $c_0(t,0)$



# Comparison of Optimal Design with Start Designs A and B

| | | Optimal design | Start design A | Start design B |
|---|---|---|---|---|
| A-Criterion | | 0.0472 | 3.8859 | 0.750 |
| $n$ | 1.0 | $\pm$ 0.0191 | $\pm$ 0.1604 | $\pm$ 0.0800 |
| $\alpha$ | 1.0 | $\pm$ 0.2438 | $\pm$ 1.5968 | $\pm$ 0.8600 |
| $K_s$ | 1.0 | $\pm$ 0.4348 | $\pm$ 3.3751 | $\pm$ 1.6735 |
| $k$ | 1.0 | $\pm$ 0.0086 | $\pm$ 0.0124 | $\pm$ 0.0083 |
| $b$ | 1.0 | $\pm$ 0.1852 | $\pm$ 3.0460 | $\pm$ 0.9712 |
| $D_m$ | 1.0 | $\pm$ 0.0108 | $\pm$ 0.2644 | $\pm$ 0.1032 |

## Influence of Profile Measurements on the Designs

| | | Experiment 1 | Experiment 2 | Experiment 3 |
|---|---|---|---|---|
| A-Criterion | | 0.0472 | 0.0749 | 0.0038 |
| $n$ | 1.0 | ± 0.0191 | ± 0.0283 | ± 0.0047 |
| $\alpha$ | 1.0 | ± 0.2438 | ± 0.2850 | ± 0.0747 |
| $K_s$ | 1.0 | ± 0.4348 | ± 0.4735 | ± 0.1126 |
| $k$ | 1.0 | ± 0.0086 | ± 0.0121 | ± 0.0001 |
| $b$ | 1.0 | ± 0.1852 | ± 0.3764 | ± 0.0691 |
| $D_m$ | 1.0 | ± 0.0108 | ± 0.0351 | ± 0.0064 |

- Experiment 1: 24 outflow measurements ($\sigma$=0.01) + 6 profile measurements ($\sigma$=0.1)
- Experiment 2: ONLY 24 outflow measurements ($\sigma$=0.01)
- Experiment 3: 24 outflow measurements ($\sigma$=0.01) + 6 profile measurements ($\sigma$=0.01)

# enario 2: Optimization of Experimental Conditions and Sampling Scheme

## Experimental Conditions

- Control functions $q(t,0)$ and $c_0(t,0)$
- Start designs as in Scenario 1

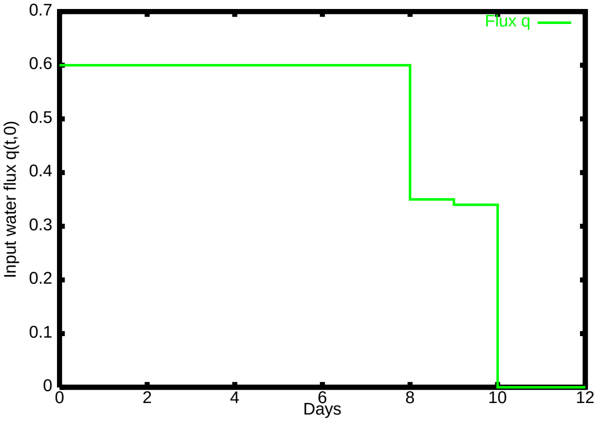
## Sampling Scheme

- Outflow Data:
  Measurements are feasible every 6 hours
  $\Longrightarrow$ Choose 24 measurements out of 48 feasible measurements
- Profile data:
  fixed as in Scenario 1

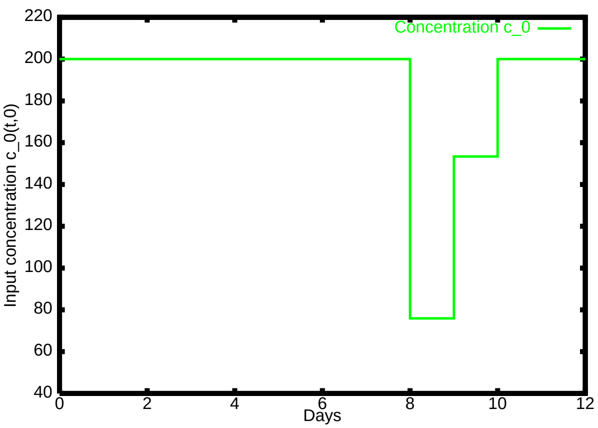
| | | Scen. 1 | Scen. 2 |
|---|---|---|---|
| A-Criterion | | 0.047 | 0.026 |
| $n$ | 1.0 | $\pm$ 0.0191 | $\pm$ 0.0215 |
| $\alpha$ | 1.0 | $\pm$ 0.2438 | $\pm$ 0.2030 |
| $K_s$ | 1.0 | $\pm$ 0.4348 | $\pm$ 0.2872 |
| $k$ | 1.0 | $\pm$ 0.00857 | $\pm$ 0.0050 |
| $b$ | 1.0 | $\pm$ 0.1852 | $\pm$ 0.1876 |
| $D_m$ | 1.0 | $\pm$ 0.0108 | $\pm$ 0.0217 |

# Optimal Experimental Conditions and Sampling Scheme

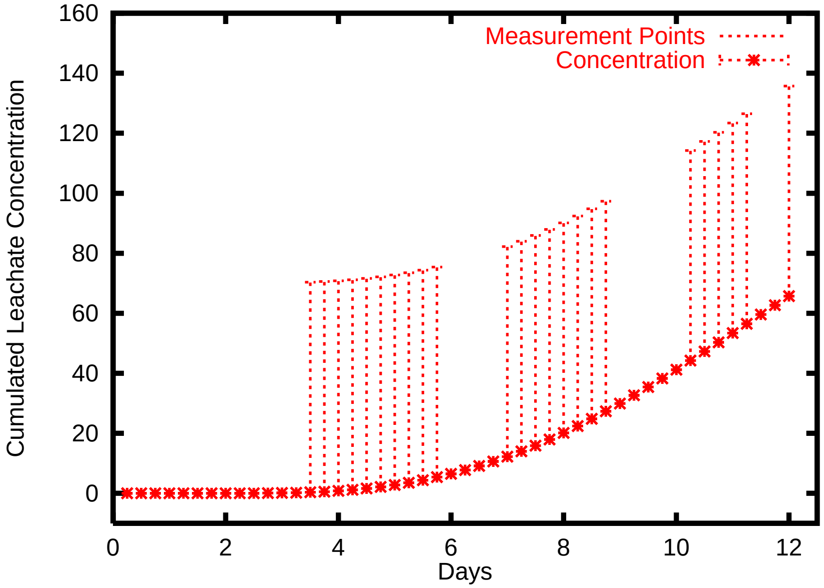
Input water flux $q(t,0)$



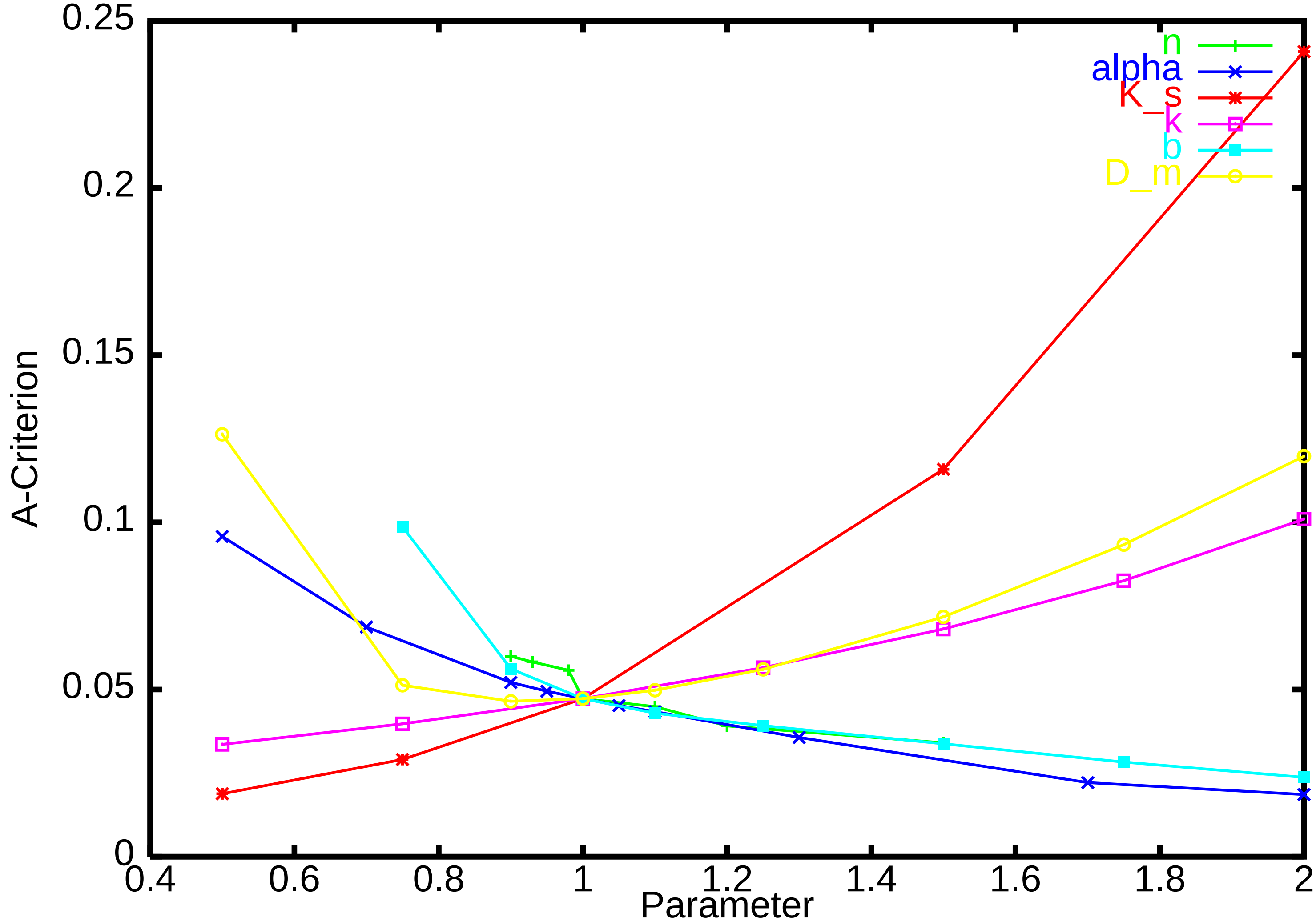

Input concentration $c_0(t,0)$



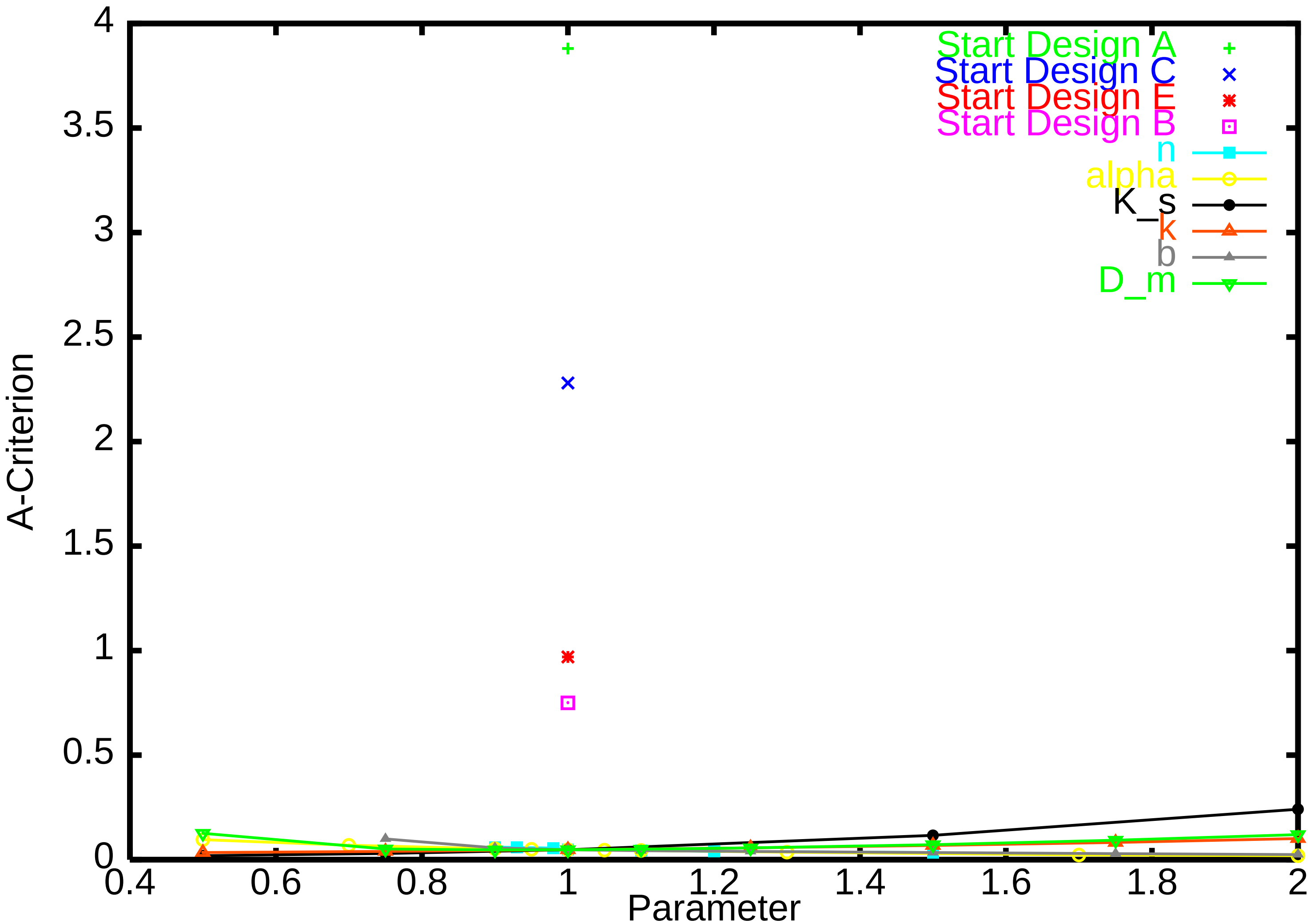

Allocation of Measurement Points



# Parameter Sensitivity of Designs



# Comparison: Start Designs - Optimal Design



# Summary and Outlook

For transport and degradation processes of pesticides in soils we developed:

- Methods for parameter estimation $\Longrightarrow$ **ECOFIT**
- Methods for optimal experimental design $\Longrightarrow$ **ECOPLAN**

- Experiments on the basis of optimized designs
- General strategies for column and lysimeter experiments $\Longrightarrow$ Registration